

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2002, 11:52:22 ; Search time 18.17 seconds

(without alignments)  
52.465 Million cell updates/sec

Title: US-09-432-546-6

Sequence: 1 RRPMPWPKWPLIGGYDPAAPPPPP 26

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69	37.7	520	1	WASP_MOUSE
2	66	36.1	10	1	TMOE_AEDAE
3	66	36.1	1171	1	DIA2_MOUSE
4	65	35.5	2715	1	TRX2_HUMAN
5	64	35.0	534	1	ABG_ARATH
6	64	35.0	833	1	DIA3_HUMAN
7	63	34.4	304	1	GAT1_CHICK
8	62	33.9	144	1	INDC_BOVIN
9	62	33.9	785	1	ISP4_SCHPO
10	61.5	33.6	445	1	AMC2_ORYSA
11	61	33.3	111	1	VPX_HV2KR
12	61	33.3	112	1	VPX_HV2CA
13	61	33.3	112	1	VPX_HV2D1
14	61	33.3	112	1	VPX_HV2RO
15	61	33.3	112	1	VPX_SIVM1
16	61	33.3	112	1	VPX_SIVM2
17	61	33.3	112	1	VPX_SIVM3
18	61	33.3	112	1	VPX_SIVM4
19	61	33.3	249	1	PRA_MYCE
20	61	33.3	1044	1	RSG1_BOVIN
21	60	32.8	825	1	FL_ORYSA
22	60	32.8	1112	1	SE5_RAT
23	60	32.8	1403	1	PRO_DROME
24	60	32.8	1403	1	CYSP_TRYBB
25	59	32.2	450	1	SR4_RAT
26	59	32.0	1048	1	SOS2_MOUSE
27	58.5	32.0	1297	1	CUT2_CABEL
28	58	31.7	231	1	PRA_MYCTU
29	58	31.7	240	1	OCT6_HUMAN
30	58	31.7	448	1	CAP_SCHPO
31	58	31.7	1059	1	CAPU_DROME
32	58	31.7	1097	1	KFID_RAT
33	58	31.7	1097	1	KFID_RAT

34	58	31.7	1139	1	RBI2_HUMAN	Q08999 homo sapien
35	57.5	31.4	112	1	VPX_HV2G1	P18045 human immun
36	57.5	31.4	112	1	VPX_HV2ST	P20881 human immun
37	57	31.1	81	1	YIDD_PSEPU	P25753 pseudomonas
38	57	31.1	400	1	SCW_DROME	P54631 drosophila
39	57	31.1	502	1	WASP_HUMAN	P42768 homo sapien
40	57	31.1	1038	1	RSG1_RAT	P50904 rattus norv
41	57	31.1	1047	1	RSG1_HUMAN	P20936 homo sapien
42	57	31.1	1101	1	DIA2_HUMAN	P060879 homo sapien
43	57	31.1	1213	1	FMN_CHICK	O05858 gallus gall
44	57	31.1	1790	1	SEPA_EMENT	P78621 emeritella
45	56.5	30.9	379	1	YB5_CABEL	Q09442 caenorhabdi

## ALIGNMENTS

RESULT	ID	WASP_MOUSE	STANDARD	PRT	520 AA.
AC	P70315	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	WISKOTT-ALDRICH SYNDROME PROTEIN HOMOLOG (WASP).				
GN	WAS OR WASP.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
NC	NCBI_TaxID=10090;				
RM	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=BAIB/C;				
RX	MEDLINE=96115600; PubMed=8666397;				
RA	Derry J.M.J., Wiedemann P., Blair P., Wang Y., Kerns J.A.,				
RA	Lemahieu V., Godfrey V.L., Wilkinson J.E., Francke U.;				
RT	"The mouse homolog of the Wiskott-Aldrich syndrome protein (WASP)				
RT	gene is highly conserved and maps near the scurfy (sf) mutation on				
RT	the x chromosome."				
RL	Genomics 29:471-477(1995).				
CC	-1- FUNCTION: POSSIBLE REGULATOR OF LYMPHOCYTE AND PLATELET FUNCTION.				
CC	MAY BE INVOLVED IN SIGNALING PATHWAYS WITH CYTOSKELETAL FUNCTION				
CC	(BY SIMILARITY).				
CC	-1- DOMAIN: THE WH1 (WASP HOMOLOG 1) DOMAIN MAY BIND A PRO-RICH				
CC	LIGAND.				
CC	-1- SIMILARITY: CONTAINS 1 GBD DOMAIN.				
CC	-1- SIMILARITY: CONTAINS 1 WH1 DOMAIN.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation-				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	-----				
DR	EMBL; U54788; AAC52556.1; -				
DR	MGI; MGI:105059; Was.				
DR	InterPro: IPR000095; PAK_box_P21_Rho_binding.				
DR	InterPro: IPR000697; RanBP1_WASP.				
DR	InterPro: IPR001960; WH1.				
DR	InterPro: IPR003124; WH2.				
DR	Pfam; PF00786; PBD; 1.				
DR	Pfam; PF00568; WH1; 1.				
DR	Pfam; PF02205; WH2; 1.				
DR	SMART; SM00285; PBD; 1.				
DR	SMART; SM00461; WH1; 1.				
DR	SMART; SM00246; WH2; 1.				
DR	PROSITE; PS50108; GBD; 1.				
KW	Repeat.				
FT	DOMAIN	41	147		WH1.
FT	DOMAIN	240	259		GBD.
FT	REPEAT	354	363		GRSGPLPPXP MOTIF 1.

FT REPEAT 393 402 GRSGPPLPPX MOTIF 2.  
 FT DOMAIN 162 167 POLY-PRO.  
 FT DOMAIN 314 321 POLY-PRO.  
 FT DOMAIN 324 341 POLY-GLY.  
 FT DOMAIN 368 373 POLY-PRO.  
 FT DOMAIN 376 379 POLY-PRO.  
 FT DOMAIN 384 390 POLY-PRO.  
 FT DOMAIN 397 403 POLY-PRO.  
 FT DOMAIN 408 424 POLY-PRO.  
 FT DOMAIN 503 520 ASP/GLU-RICH (ACIDIC).  
 FT SEQUENCE 520 AA; 54191 MW; 9C223733C59F0C8A CRC64;

Query Match 37.7%; Score 69; DB 1; Length 520;  
 Best Local Similarity 75.0%; Pred. No. 1.2;  
 Matches 12; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 11 PLIGGYDAPPPPP 26  
 DB 401 PLPGAGPPAPPPPP 416

## RESULT 2

TMOF\_AEDAE STANDARD: PRT: 10 AA.  
 ID TMOF\_AEDAE  
 AC P19425;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE TRYPSIN-MODULATING OOSTATIC FACTOR (TMOF) (OOSH).  
 OS Aedes aegypti (Yellowfever mosquito).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
 OC Aedes.  
 OX NCBI\_TaxID=7159;  
 [1]  
 RP SEQUENCE.  
 RC STRAIN-VERO BEACH; TISSUE-Ovary;  
 RA MEDLINE=90367888; PubMed=2394318;  
 RA Borovsky D., Carlson D.A., Griffin P.R., Shabanowitz J., Hunt D.F.;  
 RT "Mosquito oostatic factor: a novel decapetide modulating  
 RT trypsin-like enzyme biosynthesis in the midgut.";  
 RL FASPB J. 4:3015-3020(1990).  
 [2]  
 RP SEQUENCE.  
 RC STRAIN-VERO BEACH; TISSUE-Ovary;  
 RA MEDLINE=93357794; PubMed=833526;  
 RA Borovsky D., Carlson D.A., Griffin P.R., Shabanowitz J., Hunt D.F.;  
 RT "Mass spectrometry and characterization of Aedes aegypti trypsin  
 RT modulating oostatic factor (TMOF) and its analogs.";  
 RL Insect Biochem. Mol. Biol. 23:703-712(1993).  
 CC -1- FUNCTION: HAS AN OOSTATIC ACTIVITY. INHIBITS TRYPSIN BIOSYNTHESIS  
 CC IN THE MIDGUT WHICH INDIRECTLY REDUCES THE VITELLOGENIN  
 CC CONCENTRATION IN THE HEMOLYMPH RESULTING IN INHIBITION OF OOCYTE  
 CC DEVELOPMENT.  
 CC -1- DEVELOPMENTAL STAGE: SYNTHESIZED AND RELEASED FROM FOLLICULAR  
 CC EPITHELIUM 18-24 HRS AFTER A BLOOD MEAL. SYNTHESIS PEAKS AT  
 CC 36 HRS AND STAYS AT 56 HRS.  
 CC PIR: A36454; A36454.  
 DR Hormone.  
 FT DOMAIN 1 2 POLY-PRO.  
 FT VARIANT 1 2 YD -> DY (IN TMOF(B)).  
 FT SEQUENCE 10 AA; 1047 MW; 236DDA7777776DCT CRC64;

Query Match 36.1%; Score 66; DB 1; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.066; 0; Indels 0; Gaps 0;  
 Matches 10; Conservative 0; Mismatches 0;

OY 17 YDPAPPPPP 26  
 DB 1 YDPAPPPPP 10

RESULT 3  
 ID DIA3\_MOUSE STANDARD: PRT: 1171 AA.  
 AC Q92207;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE DIAPHANOUS PROTEIN HOMOLOG 3 (DIAPHANOUS-RELATED FORMIN 3) (DRF3)  
 DE (MDIA2) (PL3ANDIA2).  
 GN DIAPH3 OR DIA3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98204643; PubMed=9535835;  
 RA Alberts A.S., Bouquin N., Johnston L.H., Treisman R.;  
 RA Tomlinaga T., Sahai E., Chardin P., McCormick F., Courtneidge S.A.,  
 RT "Diaphanous-related formins bridge Rho GTPase and Src tyrosine kinase  
 RT signaling.";  
 RL J. Biol. Chem. 273:8616-8622(1998).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX Tomlinaga T., Sahai E., Treisman R.H., Alberts A.S.;  
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RP FUNCTION.  
 RX MEDLINE=20142655; PubMed=10678165;  
 RA Tomlinaga T., Sahai E., Chardin P., McCormick F., Courtneidge S.A.,  
 RT "Diaphanous-related formins bridge Rho GTPase and Src tyrosine kinase  
 RT signaling.";  
 RL Mol. Cell 5:13-25(2000).  
 CC -1- FUNCTION: BINDS TO GTP-BOUND FORM OF RHO AND TO PROFILIN. ACTS IN  
 CC A RHO-DEPENDENT MANNER TO RECRUIT PROFILIN TO THE MEMBRANE, WHERE  
 CC IT PROMOTES ACTIN POLYMERIZATION. IT IS REQUIRED FOR CYTOKINESIS,  
 CC STRESS FIBER FORMATION, AND TRANSCRIPTIONAL ACTIVATION OF THE  
 CC SERUM RESPONSE FACTOR. DFR PROTEINS COUPLE RHO AND SRC TYROSINE  
 CC KINASE DURING SIGNALING AND THE REGULATION OF ACTIN DYNAMICS.  
 CC -1- DOMAIN: DRES ARE REGULATED BY INTRAMOLECULAR GBD-DAD BINDING WHERE  
 CC RHO-GTP ACTIVATES THE DRES BY DISRUPTING THE GBD-DAD INTERACTION.  
 CC -1- SIMILARITY: CONTAINS 1 GTPASE-BINDING DOMAIN (GBD).  
 CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 1 (FH1) DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 2 (FH2) DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 3 (FH3) DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 DRF AUTOREGULATORY DOMAIN (DAD).  
 CC -1- SIMILARITY: BELONGS TO THE FORMIN HOMOLOG FAMILY. DIAPHANOUS  
 CC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF094519; AAC71771.1; -.  
 CC MGD: MGI:1927222; Diap3.  
 DR InterPro: IPR003104; FH2.  
 DR Pfam: PF02181; FH2; 1.  
 DR SMART: SM00498; FH2; 1.  
 DR SMART; Repeat.  
 KW Coiled coil; Repeat.  
 FT DOMAIN 81 277 GBD.  
 FT DOMAIN 176 473 FH3.  
 FT DOMAIN 373 403 COILED COIL (POTENTIAL).  
 FT DOMAIN 478 533 COILED COIL (POTENTIAL).  
 FT DOMAIN 540 610 FH1 (PRO-RICH).  
 FT DOMAIN 615 1056 FH2.  
 FT DOMAIN 887 918 COILED COIL (POTENTIAL).  
 FT DOMAIN 988 1038 COILED COIL (POTENTIAL).

FT DOMAIN 1039 1053 DAD.  
 FT DOMAIN 1055 1058 ARG/LYS-RICH (BASIC).  
 SQ SEQUENCE 1171 AA; 133685 MW; 95347A854CABC7CF CRC64;

Query Match 36.1%; Score 66; DB 1; Length 1171;  
 Best Local Similarity 62.5%; Pred. No. 5.5;  
 Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 11 PLIGGYDAPPAPP 26  
 Db 559 PALSGVPPPPPPPP 574

RESULT 4  
 TRX2\_HUMAN STANDARD; PRT: 2715 AA.  
 AC Q9UMN6; Q9UR25; Q95836; Q9Y669; Q9Y668; O15022;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DE TRITHORAX HOMOLOG 2 (MIXED LINEAGE LEUKEMIA GENE HOMOLOG 2 PROTEIN).  
 GN TRX2 OR HRX2 OR MLT2 OR KIA0340.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 CX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (LONG ISOFORM).  
 RA Angrand P.O., Valatine H., Jeannoulet F., Adamson A.,  
 RA van der Hoeven F., Olsen L., Tekotte H., Huang N., Poch O.,  
 RA Lamedin J., Chambon P., Lossen R., Stewart A., Aasland R.;  
 RA "Mammalian trithorax- and Ash1-like proteins: putative chromatin  
 RT regulators which contain PHD fingers and SET domains.";  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A. (LONG ISOFORM).  
 RA Lamedin J.E., McCreedy P.M., Adamson A.W., Burkhardt-Schultz K.,  
 RA Garcia E., Kyle A., Ramirez M., Stillwagen S., Gaines J., Danganan L.,  
 RA Bruce R., Quan G., Montgomery M., Ow D., Kobayashi A., Olsen A.O.,  
 RA Carrano A.V.;  
 RT "Sequence analysis of a 1 Mb region in human 19q13.1";  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 816-2715 FROM N.A. (LONG ISOFORM).  
 RC TISSUE-Brain;  
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,  
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. VII.  
 RL The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro.";  
 RL DNA Res. 4:141-150(1997).  
 RN [4]  
 RP SEQUENCE OF 111-2715 FROM N.A. (LONG ISOFORM).  
 RC TISSUE-Testis, and Leukocyte;  
 RA MEDLINE=20105772; PubMed=10637508;  
 RA Huntsman D.G., Chin S.-F., Mulieris M., Bailey S.J., Collins V.P.,  
 RA Wiedemann L.M., Aparicio S., Caldas C.;  
 RT "MLT2, the second human homolog of the Drosophila trithorax gene, maps  
 RL to 19q13.1 and is amplified in solid tumor cell lines.";  
 RL Oncogene 18:7975-7984(1999).  
 RN [5]  
 RP PARTIAL SEQUENCE FROM N.A. (LONG AND TRUNCATED ISOFORMS).  
 RC TISSUE-Placenta, and Bone marrow;  
 RA MEDLINE=99339983; PubMed=10409430;  
 RA Fitzgerald K.T., Diaz M.O.;  
 RT "MLT2: A new mammalian member of the trx/MLT family of genes.";  
 RL Genomics 59:187-192(1999).  
 CC -1- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A  
 CC TRUNCATED FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHEST LEVELS IN TESTIS.  
 CC ALSO FOUND IN BRAIN, BONE MARROW, HEART, MUSCLE, KIDNEY, PANCREAS,  
 CC SPLEEN, THYMUS, PROSTATE, OVARY, INTESTINE, COLON, PERIPHERAL  
 CC BLOOD LYMPHOCYTES, AND PLACENTA.  
 CC -1- DISEASE: OFTEN AMPLIFIED IN PANCREATIC CARCINOMAS.  
 CC -1- SIMILARITY: BELONGS TO THE TRITHORAX FAMILY OF TRANSCRIPTION  
 CC FACTORS.  
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 3 PHD-TYPE ZINC FINGERS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib.ch](mailto:license@sib.ch)).  
 CC -----  
 CC EMBL; AJ007041; CAB45385.1; -;  
 CC EMBL; AD000671; -; NOT\_ANNOTATED\_CDS.  
 CC EMBL; AB002302; BAA20763.2; -;  
 CC EMBL; AF186605; AAD56420.1; -;  
 CC EMBL; AF104918; AAD17932.1; -;  
 CC EMBL; AF105279; AAD26113.1; -;  
 CC EMBL; AF105280; AAD26112.1; -;  
 CC InterPro: IPR000637; AT-hook.  
 CC InterPro: IPR003889; FYRICH\_C.  
 CC InterPro: IPR003888; FYRICH\_N.  
 CC InterPro: IPR001965; PHD.  
 CC InterPro: IPR002965; P-rich\_extensn.  
 CC InterPro: IPR003616; PostSET.  
 CC InterPro: IPR001214; SET.  
 CC InterPro: IPR002857; ZnF-CXXC.  
 CC InterPro: IPR001841; ZnF-ring.  
 CC Pfam; PF00628; PHD; 3.  
 CC Pfam; PF00856; SET; 1.  
 CC Pfam; PF02008; zf-CXXC; 1.  
 CC SMART; SM00384; AT\_hook; 1.  
 CC SMART; SM00542; FYRICH; 1.  
 CC SMART; SM00541; FYRICH; 1.  
 CC SMART; SM00249; PHD; 4.  
 CC SMART; SM00508; PostSET; 1.  
 CC SMART; SM00184; RING; 1.  
 CC SMART; SM00317; SET; 1.  
 CC PROSITE; PS50280; SET; 1.  
 CC KW DNA-binding; Bromodomain; Nuclear protein; Zinc-finger; Metal-binding;  
 CC Transcription regulation; Alternative splicing.  
 CC FT DNA\_BIND 37 44  
 CC FT DNA\_BIND 110 117  
 CC FT DNA\_BIND 357 365  
 CC FT ZN\_FING 1203 1252  
 CC FT ZN\_FING 1253 1303  
 CC FT ZN\_FING 1337 1396  
 CC FT DOMAIN 1449 1471  
 CC FT DOMAIN 2586 2715  
 CC FT DOMAIN 26 37  
 CC FT DOMAIN 248 255  
 CC FT DOMAIN 362 398  
 CC FT DOMAIN 402 771  
 CC FT DOMAIN 808 812  
 CC FT DOMAIN 1963 1970  
 CC FT DOMAIN 2251 2259  
 CC FT VARSPLIC 532 582  
 CC FT VARSPLIC 583 2715  
 CC FT CONFLICT 834 834  
 CC FT CONFLICT 941 941  
 CC FT CONFLICT 1317 1317  
 CC FT CONFLICT 1362 1362  
 CC FT CONFLICT 1438 1438  
 CC -----  
 CC VARSPLIC 583 2715  
 CC CONFLICT 834 834  
 CC CONFLICT 941 941  
 CC CONFLICT 1317 1317  
 CC CONFLICT 1362 1362  
 CC CONFLICT 1438 1438  
 CC -----  
 CC MISSING (IN ISOFORM TRUNCATED).  
 CC K -> E (IN REF. 5).  
 CC S -> Y (IN REF. 5).  
 CC E -> Q (IN REF. 5).  
 CC H -> Y (IN REF. 5).  
 CC D -> N (IN REF. 5).

FT CONFLICT 2622 2622 D -> H (IN REF. 5)  
SQ SEQUENCE 2715 AA; 293511 MW; C0615B981BBE7BF CRC64;

Query Match 35.5%; Score 65; DB 1; Length 2715;  
Best Local Similarity 45.5%; Pred. No. 15;  
Matches 15; Conservative 1; Mismatches 7; Indels 10; Gaps 2;

Oy 4 PWMKMKW-----PLIG-----GGYDAPAPPPPP 26  
Db 2226 PLAPTSTWLPPLGLVLPVGVAPAPPPPP 2258

RESULT 5  
APG\_ARATH STANDARD; PRT; 534 AA.

AC PA0602;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE ANTER-SPECIFIC PROLINE-RICH PROTEIN APG PRECURSOR.  
GN APC.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eustoids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94004980; PubMed=8401599;  
RA Roberts M.R., Foster G.D., Blundell R.P., Robinson S.W., Kumar A.,  
RA Draper J., Scott R.;  
RT "Gemeophytic and sporophytic expression of an anther-specific  
Arabidopsis thaliana gene.";  
RL Plant J. 3:111-120(1993).  
CC -1- TISSUE SPECIFICITY: FOUND IN SPOROPHYTIC AND GEMETOPHYTIC CELL  
CC TYPES IN THE ANTER, ONLY IN MALE FERTILE PLANTS.  
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN MALE GEMETOGENESIS, DURING  
CC MICROSPORE DEVELOPMENT. HIGHER EXPRESSION IS FOUND DURING  
CC MICROSPORE MITOSIS WITH A DRAMATIC DECLINE DURING POLLEN  
CC MATURATION.  
CC -1- SIMILARITY: BELONGS TO THE "GDSL" FAMILY OF LIPOLYTIC ENZYMES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X60377; CAA42925.1; -  
DR PIR; S21961; S21961.  
DR InterPro; IPR001087; Lipase\_GDSL.  
DR Pfam; PF00657; Lipase\_GDSL; 1.  
DR PROSITE; PS01098; LIPASE\_GDSL\_SER; 1.  
KW Signal.  
FT SIGNAL 1 35 POTENTIAL.  
FT CHAIN 36 534 ANTER-SPECIFIC PROLINE-RICH PROTEIN APG.  
FT ACT\_SITE 211 211 BY SIMILARITY.  
FT ACT\_SITE 511 511 POTENTIAL.  
SQ SEQUENCE 534 AA; 57967 MW; 744CAD3B08CC482E CRC64;

Query Match 35.0%; Score 64; DB 1; Length 534;  
Best Local Similarity 37.5%; Pred. No. 4.3;  
Matches 12; Conservative 3; Mismatches 11; Indels 6; Gaps 1;

Oy 1 RRPW-----WPKWPLIGGYDAPAPPPPP 26  
Db 41 RLPWPLWMPRPYPPQPMKMPPPDPSPKRVAP 72

RESULT 6  
DIA3\_HUMAN STANDARD; PRT; 853 AA.  
ID DIA3\_HUMAN  
AC Q9NSV4;  
DT 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE DIAPHANOUS PROTEIN HOMOLOG 3 (DIAPHANOUS-RELATED FORMIN 3) (DRF3)  
(FRAGMENT)  
GN DIAPH3 OR DIAP3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RC TISSUE=Testis;  
RA Blum H., Bauersachs S., Mewes H.-W., Gassenhuber J., Wiemann S.;  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 468-853 FROM N.A. (ISOFORMS 1 AND 2).  
RA Smith M.;  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: BINDS TO GTP-BOUND FORM OF RHO AND TO PROFILIN. ACTS IN  
CC A RHO-DEPENDENT MANNER TO RECRUIT PROFILIN TO THE MEMBRANE, WHERE  
CC IT PROMOTES ACTIN POLYMERIZATION. IT IS REQUIRED FOR CYTOKINESIS,  
CC STRESS FIBER FORMATION, AND TRANSCRIPTIONAL ACTIVATION OF THE  
CC SERUM RESPONSE FACTOR. DFR PROTEINS COUPLE RHO AND SRC TYROSINE  
CC KINASE DURING SIGNALING AND THE REGULATION OF ACTIN DYNAMICS (BY  
CC SIMILARITY).  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE  
CC PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- DOMAIN: DRFS ARE REGULATED BY INTRAMOLECULAR GBD-DAD BINDING WHERE  
CC RHO-GTP ACTIVATES THE DRFS BY DISRUPTING THE GBD-DAD INTERACTION  
CC (BY SIMILARITY)  
CC -1- SIMILARITY: CONTAINS 1 GTPASE-BINDING DOMAIN (GBD).  
CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 1 (FH1) DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 2 (FH2) DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 3 (FH3) DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 DRF AUTOREGULATORY DOMAIN (DAD).  
CC -1- SIMILARITY: BELONGS TO THE FORMIN HOMOLOG FAMILY. DIAPHANOUS  
CC SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AL137718; CAB70890.1; ALT\_INT.  
DR EMBL; AL354829; CAC17664.1; -  
DR EMBL; AL354829; CAC17665.1; -  
DR InterPro; IPR002205; DNA\_topoisolv.  
DR InterPro; IPR003104; FH2.  
DR InterPro; IPR003109; Golooco.  
DR Pfam; PF00521; DNA\_topoisolv; 1.  
DR Pfam; PF02181; FH2; 1.  
DR Pfam; PF02188; Golooco; 1.  
DR SMART; SM00498; FH2; 1.  
KW Coiled coil; Repeat; Alternative splicing.  
FT NON\_TER 1 1  
FT DOMAIN 304 374 FHL (PRO-RICH).  
FT DOMAIN 379 818 FH2.  
FT DOMAIN 801 815 DAD.  
FT DOMAIN 137 167 COILED COIL (POTENTIAL).  
FT DOMAIN 241 299 COILED COIL (POTENTIAL).  
FT DOMAIN 650 799 COILED COIL (POTENTIAL).  
FT DOMAIN 819 822 ARG/LYS-RICH (BASIC).  
FT DOMAIN 819 822 VSETELEKRWQGRQDLEKELETFFPPEDHDKFYTKI  
FT VSETELEKRWQGRQDLEKELETFFPPEDHDKFYTKI  
FT F -> GLGFKHFKMALIFSARKLITPTICWTFPLSHSVF  
FT IPNISF (IN ISOFORM 2).  
FT

FT VARSPLIC 698 853 MISSING (IN ISOFORM 2).  
 SQ SEQUENCE 853 AA; 98610 MW; B7FA9C745AE18CD9 CRC64;

Query Match 35.0%; Score 64; DB 1; Length 853;  
 Best Local Similarity 76.9%; Pred. No. 6.6;  
 Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 14 GGGYDPAPPPPP 26  
 ||| | |||||  
 Db 327 GGGYDPAPPPPP 339

RESULT 7  
 GATL CHICK  
 ID GATL CHICK STANDARD; PRT; 304 AA.

AC p1678:  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE ERYTHROID TRANSCRIPTION FACTOR (GATA-1) (ERYFL) (NF-E1 DNA-BINDING  
 DE PROTEIN) (NF-E1A).  
 GN GATA1 OR ERYFL.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauvia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN 111  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89376538; PubMed=2776214;  
 RA Evans T., Felsenfeld G.;  
 RT "The erythroid-specific transcription factor Eryfl: a new finger  
 RT protein.";  
 RL Cell 58:877-885(1989).  
 RN 121  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91065513; PubMed=2249770;  
 RA Yamamoto M., Ko L.J., Leonardo M.W., Beug H., Orkin S.H.,  
 Engel J.D.;  
 RT "Activity and tissue-specific expression of the transcription factor  
 RT NF-E1 multigene family.";  
 RL Genes Dev. 4:1650-1662(1990).  
 RN 131  
 RP STRUCTURE BY NMR OF 158-223.  
 RX MEDLINE=93324913; PubMed=8332909;  
 RA Omichinski J.G., Clore G.M., Schaad O., Felsenfeld G., Trainor C.,  
 RA Appella E., Stahl S.J., Gronenborn A.M.;  
 RT "NMR structure of a specific DNA complex of Zn-containing DNA binding  
 RT domain of GATA-1.";  
 RL Science 261:438-446(1993).  
 RN 141  
 RP STRUCTURE BY NMR OF 158-223.  
 RX MEDLINE=97448676; PubMed=9303001;  
 RA Tjandra N., Omichinski J.G., Gronenborn A.M., Clore G.M., Bax A.;  
 RT "Use of dipolar 1H-15N and 1H-13C couplings in the structure  
 RT determination of magnetically oriented macromolecules in solution.";  
 RL Nat. Struct. Biol. 4:732-738(1997)  
 CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR WHICH PROBABLY SERVES AS A  
 CC GENERAL SWITCH FACTOR FOR ERYTHROID DEVELOPMENT. IT BINDS TO DNA  
 CC SITES WITH THE CONSENSUS SEQUENCE (A/T)GATA(A/G) WITHIN REGULATORY  
 CC REGIONS OF GLOBIN GENES AND OF OTHER GENES EXPRESSED IN ERYTHROID  
 CC CELLS.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- TISSUE SPECIFICITY: ERYTHROCYTE.  
 CC -1- DOMAIN: THE TWO FINGERS ARE FUNCTIONALLY DISTINCT AND COOPERATE TO  
 CC ACHIEVE SPECIFIC, STABLE DNA BINDING. THE FIRST FINGER IS  
 CC NECESSARY ONLY FOR FULL SPECIFICITY AND STABILITY OF BINDING.  
 CC WHEREAS THE SECOND ONE IS REQUIRED FOR BINDING (BY SIMILARITY).  
 CC -1- SIMILARITY: CONTAINS 2 GATA-TYPE ZINC FINGERS.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----  
 DR EMBL; M26209; AAA49055.1; .  
 DR PIR; A32993; A32993.  
 DR PDB; 1GAT; 31-OCT-93.  
 DR PDB; 2GAT; 28-JAN-98.  
 DR PDB; 3GAT; 28-JAN-98.  
 DR PDB; 1GAU; 31-OCT-93.  
 DR TRANSFAC; T00267; .  
 DR InterPro; IPR002965; P-rich\_extensions.  
 DR InterPro; IPR000679; ZnF\_GATA.  
 DR InterPro; IPR001164; ZnF\_GCS.  
 DR Pfam; PF00320; GATA; 2.  
 DR PRINTS; PR00619; GATAZNFINGER.  
 DR SMART; SM00401; ZnF\_GATA; 2.  
 DR PROSITE; PS00344; GATA\_ZN\_FINGER\_1; 2.  
 DR PROSITE; PS50114; GATA\_ZN\_FINGER\_2; 2.  
 KW Transcription regulation; Activator; DNA-binding; Erythrocyte;  
 KW Zinc-finger; Nuclear protein; 3D-structure.  
 FT ZN\_FING 110 134  
 FT ZN\_FING 110 188  
 FT ZN\_FING 164 161  
 FT TURN 160 167  
 FT TURN 165 167  
 FT STRAND 175 178  
 FT TURN 179 181  
 FT STRAND 182 185  
 FT HELIX 186 195  
 FT TURN 196 196  
 FT TURN 202 203  
 SQ SEQUENCE 304 AA; 31417 MW; 64C9D6FDB58CE83F CRC64;

Query Match 34.4%; Score 63; DB 1; Length 304;  
 Best Local Similarity 71.4%; Pred. No. 3.3;  
 Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 13 IGGYDPAPPPPP 26  
 :||| ||: |||||  
 Db 240 MGGGDPSPMPPPP 253

RESULT 8  
 INDC\_BOVIN  
 ID INDC\_BOVIN STANDARD; PRT; 144 AA.  
 AC P33046;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE INDOLICIDIN PRECURSOR.  
 OS Bos taurus (Bovine).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 CC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN 111  
 RP SEQUENCE FROM N.A.  
 RP TISSUE-BONE MARROW.  
 RX MEDLINE=92392368; PubMed=1520337;  
 RA del Sal G., Storici P., Schneider C., Romeo D., Zanetti M.;  
 RT "cDNA cloning of the neutrophil bactericidal peptide indolicidin.";  
 RL Biochem. Biophys. Res. Commun. 187:467-472(1992).  
 RN 121  
 RP SEQUENCE OF 131-143.  
 RP TISSUE=Neutrophils;  
 RX MEDLINE=92165771; PubMed=1537821;  
 RA Seisted M.E., Novotny M.J., Morris W.L., Tang Y.-Q., Smith W.,  
 RA Cullor J.S.;  
 RT "Indolicidin, a novel bactericidal tridecapeptide amide from  
 RT neutrophils.";

RL J. Biol. Chem. 267:4292-4295(1992).  
 CC -1- FUNCTION: POTENT MICROBICIDAL ACTIVITY. ACTIVE AGAINST  
 CC STAPHYLOCOCCUS AUREUS AND ESCHERICHIA COLI.  
 CC -1- TISSUE SPECIFICITY: LARGE GRANULES OF NEUTROPHILS.  
 CC -1- PTM: ELASTASE MIGHT BE RESPONSIBLE FOR ITS MATURATION.  
 CC -1- SIMILARITY: BELONGS TO THE CATHELICIDIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X67340; CAA47755.1; -  
 CC DR PIR: JCI222; JCI222.  
 CC DR PIR: A42387; A42387.  
 CC DR InterPro: IPR001894; Cathelicidin.  
 CC Pfam: PF00666; Cathelicidins\_1.  
 CC DR ProDom: PD001838; Cathelicidin\_1.  
 CC DR PROSITE: PS00946; CATHELICIDINS\_1; 1.  
 CC DR PROSITE: PS00947; CATHELICIDINS\_2; 1.  
 CC FT Antibiatic; Amidation; signal.  
 CC FT SIGNAL 1 29 POTENTIAL.  
 CC FT PROPEP 30 130  
 CC FT PEPTIDE 131 143 INDOLICIDIN.  
 CC FT MOD\_RES 30 30 PYRROLIDONE CARBOXYLIC ACID (BY  
 CC SIMILARITY).  
 CC FT DISULFID 85 96 BY SIMILARITY.  
 CC FT DISULFID 107 124 BY SIMILARITY.  
 CC FT MOD\_RES 143 143 AMIDATION (G-144 PROVIDE AMIDE GROUP).  
 CC SQ SEQUENCE 144 AA; 16479 MW; E3B1CB8E55C09911 CRC64;  
  
 Query Match 33.9%; Score 62; DB 1; Length 144;  
 Best Local Similarity 75.0%; Pred. No. 2.1;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 2 RWPMPMK 9  
 Db 135 KMPMPWR 142  
  
 RESULT 9  
 ISP4\_SCHPO STANDARD; PRT; 785 AA.  
 ID ISP4\_SCHPO  
 AC P40900; Q9HGP2; P78943;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE SEXUAL DIFFERENTIATION PROCESS PROTEIN ISP4.  
 GN ISP4 OR SPEPC29B5.02C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Schizosaccharomycetes; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 CC Schizosaccharomycetes.  
 CC NCBL\_TaxID=4896;  
 CC [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-95042833; PubMed-7954893;  
 RA Sato S., Suzuki H., Widyastuti U., Hotta Y., Tabata S.;  
 RT "Identification and characterization of genes induced during sexual  
 RT differentiation in Schizosaccharomyces pombe";  
 RL Curr. genet. 26:31-37(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Kohnosu A., Niwa O., Yano M., Saitoh S., Katayama T., Nagao K.,  
 RA Yanojida M.;  
 RT "S.pombe chromosome II cosmid 1228 sequence";  
 RL Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.  
 RN [3]

RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Wood V., Rajandream M.A., Barrell B.G., Saunders D., Harris D.;  
 RL Submitted (AUG-1999) to the EMBL/Genbank/DBJ databases.  
 CC -1- DEVELOPMENTAL STAGE: TRANSCRIBED SPECIFICALLY DURING SEXUAL  
 CC DEVELOPMENT.  
 CC -1- CAUTION: REF.2 SEQUENCE IS INCORRECT DUE TO FRAMESHIFTS AND OTHER  
 CC SEQUENCING ERRORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D14061; BAA03147.1; -  
 CC DR EMBL: D83992; BAA12193.1; ALT\_FRAME.  
 CC DR EMBL: AL391603; CAC05511.1; -  
 CC DR PIR: S43741; S43741.  
 CC DR PIR: S45495; S45495.  
 CC FT CONFLICT 725 725 MISSING (IN REF. 1).  
 CC FT CONFLICT 777 785 GEEFGPAEW -> R (IN REF. 1).  
 CC SQ SEQUENCE 785 AA; 89192 MW; 3F19825269BFA75D CRC64;  
  
 Query Match 33.9%; Score 62; DB 1; Length 785;  
 Best Local Similarity 44.8%; Pred. No. 10;  
 Matches 13; Conservative 0; Mismatches 10; Indels 6; Gaps 1;  
  
 QY 3 WPMMPMK-----PLGGGYDPAPPPP 25  
 Db 662 WKMPQKMGQGNPLFGGTGYTPATP 690  
  
 RESULT 10  
 AMC2\_ORYSA STANDARD; PRT; 445 AA.  
 ID AMC2\_ORYSA  
 AC P27941;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 01-DEC-1992 (Rel. 24, Last annotation update)  
 DE ALPHA-AMYLASE ISOZYME C2 PRECURSOR (EC 3.2.1.1) (1,4-ALPHA-D-GLUCAN  
 DE GLUCANOHYDROLASE).  
 GN AMY1.8.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 CC NCBL\_TaxID=4530;  
 CC [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, IR26; TISSUE=Seed;  
 RA Goldman S., Mawal Y., Wu R.;  
 RL Submitted (FEB-1992) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: IMPORTANT FOR BREAKDOWN OF ENDOSPERM STARCH DURING  
 CC GERMINATION.  
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-ALPHA-D-GLUCOSIDIC  
 CC LINKAGES IN OLIGOSACCHARIDES AND POLYSACCHARIDES.  
 CC -1- COFACTOR: BINDS A CALCIUM ION REQUIRED FOR ITS ACTIVITY.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED AT A HIGH LEVEL DURING GERMINATION  
 CC IN THE ALEURONS CELLS UNDER THE CONTROL OF THE PLANT HORMONE  
 CC GIBBERELLIC ACID AND IN THE DEVELOPING GRAINS AT A LOW LEVEL.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO  
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X64619; CAA45903.1; -  
DR PIR: S19990; S19990.  
DR HSSP: P04063; 1AMY.  
DR Mendel; 9695; ORISA; Amy1.8.  
DR InterPro: IPR000461; Alpha-amylase.  
DR Pfam: PF00128; alpha-amylase; 1.  
DR PRINTS: PR00110; ALPHAMYLASE.  
KW Hydrolase; Glycosidase; Carbohydrate metabolism; Calcium; Signal;  
KW Multigene family.  
FT SIGNAL 1 ?  
FT CHAIN 1 ?  
FT ACT\_SITE 202 445 ALPHA-AMYLASE ISOZYME C2.  
FT ACT\_SITE 314 314 BY SIMILARITY.  
FT CA\_BIND 113 113 BY SIMILARITY.  
FT CA\_BIND 172 172 BY SIMILARITY.  
SQ SEQUENCE 445 AA; 49207 MW; DED23701E836ACDA CRC64;

Query Match 33.6%; Score 61.5; DB 1; Length 445;  
Best Local Similarity 42.3%; Pred. No. 6.7;  
Matches 11; Conservative 2; Mismatches 6; Indels 7; Gaps 1;

OY 1 RRPWWPKWPLIGGYDPAPPPPP 26  
DB 416 RPPWYAVE-----KTPPPPPPP 434

## RESULT 11

VPX\_HV2KR STANDARD; PRT; 111 AA.  
AC Q74122;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE VPX PROTEIN (X ORF PROTEIN) (VIRAL ACCESSORY PROTEIN).  
GN VPX.  
OS Human immunodeficiency virus type 2 (isolate KR) (HIV-2).  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=73484;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kraus G.K., Talbot R., Leavitt M., Luznick L., Schmidt A.,  
RA Badel P., Baritz C., Morton W., Wong-Staal F., Looney D.J.,  
RL Submitted (APR-1995) to the EMBL/Genbank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: U22047; AAA64578.1; -  
DR InterPro: IPR000012; HIV\_ORFXR.  
DR Pfam: PF00522; VPR; 1.  
DR PRINTS: PR00444; HIVVPRVX.  
KW AIDS.  
SQ SEQUENCE 111 AA; 12619 MW; 00B36C204C0C364E CRC64;

Query Match 33.3%; Score 61; DB 1; Length 111;  
Best Local Similarity 75.0%; Pred. No. 2.1;  
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
OY 15 GGYDPAPPPPP 26  
DB 97 GGMRRGPPPPPP 108

## RESULT 12

VPX\_HV2CA STANDARD; PRT; 112 AA.  
AC P24110;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE VPX PROTEIN (X ORF PROTEIN) (VIRAL ACCESSORY PROTEIN).  
GN VPX.  
OS Human immunodeficiency virus type 2 (isolate CAM2) (HIV-2).  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11715;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-9117059; PubMed-2005437;  
RA Tristram M., Hill F., Karpas A.;  
RT "Nucleotide sequence of a Guinea-Bissau-derived human  
RT immunodeficiency virus type 2 proviral clone (HIV-2CAM2).";  
RT J. Gen. Virol. 72:721-724(1991).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: D00835; BAA00712.1; -  
DR PIR: D38475; ASLJGX.  
DR InterPro: IPR000012; HIV\_ORFXR.  
DR Pfam: PF00522; VPR; 1.  
DR PRINTS: PR00444; HIVVPRVX.  
KW AIDS.  
SQ SEQUENCE 112 AA; 12820 MW; A717971725B94A7E CRC64;

Query Match 33.3%; Score 61; DB 1; Length 112;  
Best Local Similarity 75.0%; Pred. No. 2.1;  
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 15 GGYDPAPPPPP 26  
DB 97 GGMRRGPPPPPP 108

## RESULT 13

VPX\_HV2D1 STANDARD; PRT; 112 AA.  
AC P17760;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE VPX PROTEIN (X ORF PROTEIN) (VIRAL ACCESSORY PROTEIN).  
GN VPX.  
OS Human immunodeficiency virus type 2 (isolate D194) (HIV-2).  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11713;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-89184631; PubMed-2467304;  
RA Kuehn H., von Briesen H., Dietrich U., Adamaki M., Mix D.,  
RA Blesert L., Kreutz R., Immelman A., Henco K., Melchner C.,  
RA Andreessen R., Gelderblom H., Ruebsamen-Waigmann H.;  
RT "Molecular cloning of two west African human immunodeficiency virus  
RT type 2 isolates that replicate well in macrophages: a Gambian  
RT isolate, from a patient with neurologic acquired immunodeficiency  
RT syndrome, and a highly divergent Ghanaian isolate.";  
RT Proc. Natl. Acad. Sci. U.S.A. 86:2383-2387(1989).  
RL [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-91045094; PubMed-2235509;  
RA Kuehn H., Kreutz R., Ruebsamen-Waigmann H.;



RT 'Nucleotide sequence of HIV-2D194, an isolate from a Gambian case of  
 RT 'neuro-AIDS', which showed excellent growth in macrophages.';  
 RL Nucleic Acids Res. 18:6142-6142(1990).  
 CC -1- MISCELLANEOUS: THIS ISOLATE IS FROM A GAMBIAN CASE OF  
 CC 'NEURO-AIDS'.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: J04542; AAA76843.1; -  
 DR EMBL: X52223; CAA36467.1; -  
 DR PIR: S12155; S12155.  
 DR HIV: J04542; VPS2D194.  
 DR InterPro: IPR00012; HIV\_ORFXR.  
 DR Pfam: PF00522; VPR; 1.  
 DR PRINTS: PR00444; HIVPRVPX.  
 KW AIDS.  
 SQ SEQUENCE 112 AA; 12820 MW; 747AC8D908EA66B4 CRC64;  
 SO  
 Query Match 33.3%; Score 61; DB 1; Length 112;  
 Best Local Similarity 75.0%; Pred. No. 2.1;  
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 15 GGYDPAPPPPP 26  
 DB 97 GGMRRGPPPPPP 108  
 ID VPS\_HY2RO STANDARD; PRT; 112 AA.  
 AC P06939;  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE VPX PROTEIN (X ORF PROTEIN) (VIRAL ACCESSORY PROTEIN).  
 OS Human immunodeficiency virus type 2 (isolate ROD) (HIV-2).  
 GN Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OC NCBI\_TaxID=11720;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=87173056; PubMed=3031510;  
 RA Guyader M., Emerman M., Sonigo P., Clavel F., Montagnier L.,  
 RA Allzon M.;  
 RT "Genome organization and transactivation of the human  
 RT immunodeficiency virus type 2".  
 RL Nature 326:662-669(1987).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: M15390; AAB00766.1; -  
 DR EMBL: X05291; -; NOT\_ANNOTATED\_CDS.  
 DR PIR: I26262; ASLJX2.  
 DR HIV: M15390; VPS2ROD.  
 DR InterPro: IPR00012; HIV\_ORFXR.  
 DR Pfam: PF00522; VPR; 1.  
 DR PRINTS: PR00444; HIVPRVPX.  
 KW AIDS.  
 SQ SEQUENCE 112 AA; 12815 MW; 0A677EB6BDB5F665 CRC64;  
 SO

Query Match 33.3%; Score 61; DB 1; Length 112;  
 Best Local Similarity 75.0%; Pred. No. 2.1;  
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 15 GGYDPAPPPPP 26  
 DB 97 GGMRRGPPPPPP 108  
 ID VPS\_STV1 STANDARD; PRT; 112 AA.  
 AC P05917;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 01-OCT-1988 (Rel. 12, Last annotation update)  
 DE VPX PROTEIN (X ORF PROTEIN).  
 OS Simian immunodeficiency virus (Mm142-83 isolate) (STV-MAC).  
 GN Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OC NCBI\_TaxID=11733;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=87287230; PubMed=3649576;  
 RA Chakrabarti L., Guyader M., Allzon M., Daniel M.D., Desrosiers R.C.,  
 RA Tiollais P., Sonigo P.;  
 RT "Sequence of simian immunodeficiency virus from macaque and its  
 RT relationship to other human and simian retroviruses".  
 RL Nature 328:543-547(1987).  
 CC -1- MISCELLANEOUS: THIS IS A MACAQUE ISOLATE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: Y00277; CAA68382.1; -  
 DR PIR: D28887; ASLJX3.  
 DR HIV: M16403; VPSXMM142.  
 DR InterPro: IPR00012; HIV\_ORFXR.  
 DR Pfam: PF00522; VPR; 1.  
 DR PRINTS: PR00444; HIVPRVPX.  
 KW AIDS.  
 SQ SEQUENCE 112 AA; 12906 MW; 46F564F45AFD960 CRC64;  
 SO

Query Match 33.3%; Score 61; DB 1; Length 112;  
 Best Local Similarity 75.0%; Pred. No. 2.1;  
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 15 GGYDPAPPPPP 26  
 DB 97 GGMRRGPPPPPP 108  
 ID VPS\_HY2RO STANDARD; PRT; 112 AA.  
 AC P06939;  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE VPX PROTEIN (X ORF PROTEIN) (VIRAL ACCESSORY PROTEIN).  
 OS Human immunodeficiency virus type 2 (isolate ROD) (HIV-2).  
 GN Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OC NCBI\_TaxID=11720;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=87173056; PubMed=3031510;  
 RA Guyader M., Emerman M., Sonigo P., Clavel F., Montagnier L.,  
 RA Allzon M.;  
 RT "Genome organization and transactivation of the human  
 RT immunodeficiency virus type 2".  
 RL Nature 326:662-669(1987).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: M15390; AAB00766.1; -  
 DR EMBL: X05291; -; NOT\_ANNOTATED\_CDS.  
 DR PIR: I26262; ASLJX2.  
 DR HIV: M15390; VPS2ROD.  
 DR InterPro: IPR00012; HIV\_ORFXR.  
 DR Pfam: PF00522; VPR; 1.  
 DR PRINTS: PR00444; HIVPRVPX.  
 KW AIDS.  
 SQ SEQUENCE 112 AA; 12815 MW; 0A677EB6BDB5F665 CRC64;  
 SO

Search completed: January 30, 2002, 11:52:23  
 Job time: 202 sec





